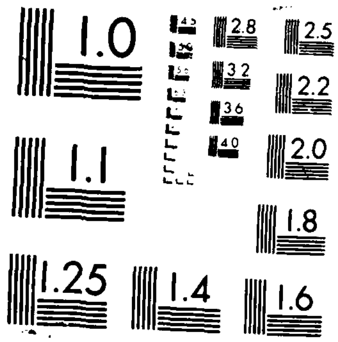
AD-A186 292 RECURSIVE M-ESTIMATORS OF LOCATION AND SCALE FOR DEPENDENT SEQUENCES(U) NORTH CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHASTIC PROC J ENGLUND ET AL. 1/1

UNCLASSIFIED NOV 86 AFOSR-TR-87-1251 F49620-85-C-0144 F/G 12/3 NL



ROCOPY RESOLUTION TEST CHART

DTIC FILE COPY

AD-A186 292

REPORT DOCUMENTATION PAGE

DTIC ELECTE OCT 1 4 1987 S D

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY: NA

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE: NA

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for Public Release; Distribution Unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): Unnumbered

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 87-1251

6a. NAME OF PERFORMING ORGANIZATION: University of North Carolina

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM

6c. ADDRESS (City, State and ZIP Code): Center for Stochastic Processes, Statistics Department, Philiips Hall 039-A, Chapel Hill, NC 27514

7b. ADDRESS (City, State and ZIP Code): Bldg. 410 Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL (If applicable): NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F49620 85 C 0144

8c. ADDRESS (City, State and ZIP Code): Bldg. 410 Bolling AFB, DC

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
6.1102F	2304	A5	

11. TITLE (Include Security Classification): Recursive M-estimators of location and scale for dependent sequences

12. PERSONAL AUTHOR(S): Englund, J.-E., Holst, U. and Ruppert, D.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
preprint	FROM 10/1/86 TO 9/30/87	November 1986	18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.
XXXXX	XXXXXXXX	XXXX

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Keywords: Robust estimation, recursive M-estimators, stochastic approximation, strong mixing, strong convergence

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Recursive M-estimators of location and scale may be obtained via stochastic approximation algorithms. We consider the case when the observations can be described by a strictly stationary process satisfying certain strong mixing conditions and results on strong convergence are given. The asymptotic distributions of the estimators for sequences of independent observations are also discussed.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: ~~Peggy Ravitch~~ Maj Crowley

22b. TELEPHONE NUMBER (Include Area Code): ~~919-962-2307~~ 202 767-5025

22c. OFFICE SYMBOL: AFOSR/NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE. UNCLASSIFIED

CODEN: LUNFD6(NFMS-3096)/1-18/(1986)

Department of Mathematical Statistics
Box 118
S-221 00 Lund, Sweden

AFOSR-TR- 87-1251

Recursive M-estimators of location and scale for dependent sequences

by

Jan-Eric Englund, Ulla Holst
and David Ruppert *

1986-9

November 1986

* The University of North Carolina
at Chapel Hill

TYP AV DOKUMENT			DOKUMENTBETECKNING/CODEN
☐ Doktorsavhandling	☐ Ansökan	☐ Tidskriftsartikel	LUNFD6(NFMS-3096)/1-18/(1986)
☐ Examensarbete	☐ Reserapport	☐ Konferensuppsats	
☐ Kompendium	☐ Delrapport	☒ Rapport.........	
	☐ Slutrapport		

AVDELNING/INSTITUTION

Department of Mathematical Statistics
Box 118, S-221 00 Lund, Sweden

FÖRFATTARE

Jan-Eric Englund, Ulla Holst and David Ruppert

DOKUMENTTITEL OCH UNDERTITEL

Recursive M-estimators of location and scale for dependent sequences

SAMMANFATTNING

Recursive M-estimators of location and scale may be obtained via stochastic approximation algorithms. We consider the case when the observations can be described by a strictly stationary process satisfying certain strong mixing conditions and results on strong convergence are given. The asymptotic distributions of the estimators for sequences of independent observations are also discussed.

NYCKELORD Robust estimation, recursive M-estimators, stochastic approximation, strong mixing, strong convergence

DOKUMENTTITEL OCH UNDERTITEL – SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL

TILLÄMPNINGSOMRÅDE

TILLÄMPNINGSNYCKELORD

UTGIVNINGSDATUM		ANTAL SID	SPRÅK
år 86	mån 11	18	☐ svenska ☒ engelska ☐ annat

ÖVRIGA BIBLIOGRAFISKA UPPGIFTER

Supported in part by Air Force Office of Scientific Research Contract AFOSR-S-49620-85-C-0144 and Swedish Natural Science Research Council F 3741-103

ISSN

ISBN

PRIS

6

1. INTRODUCTION.

P. Huber introduced the simultaneous M-estimates of location and scale, η and σ, based on observations $y_1,\ldots\ldots,y_n$, as a solution (T_n,S_n) of

$$(1.1)\qquad \begin{cases} \sum_{i=1}^{n} \psi(S_n^{-1}(y_i-T_n)) = 0, \\ \sum_{i=1}^{n} \chi(S_n^{-1}(y_i-T_n)) = 0, \end{cases}$$

where ψ and χ are suitably chosen functions. In most cases ψ is an odd and χ an even function. In particular he studied M-estimators generated by functions ψ and χ of the form (Huber's Proposal 2)

$$(1.2)\qquad \begin{cases} \psi(x) = \mathrm{sign}(x)\min(|x|,k), \\ \chi(x) = \min(k^2,x^2) - \beta_k, \end{cases}$$

with β_k chosen to make $E(\chi(z)) = 0$ if the distribution of z is $N(0,1)$. We refer to the books by Huber (1981) or Hampel et al (1986) for a review of the properties of the M-estimators. There it is proved that, if the observations are i.i.d. with a symmetric distribution, ψ is an odd and χ an even function, it follows that

$$(1.3)\qquad \begin{cases} n^{\frac{1}{2}}(T_n-\eta) \in AsN(0, \sigma^2E(\psi^2(z_1))/(E(\psi'(z_1)))^2), \\ n^{\frac{1}{2}}(S_n-\sigma) \in AsN(0, \sigma^2E(\chi^2(z_1))/(E(z_1\chi'(z_1)))^2), \end{cases}$$

where $z_1 = \sigma^{-1}(y_1-\eta)$. In this case T_n and S_n are asymptotically independent.

In real time situations, where the estimate is updated when new observations are obtained, it is often preferable to use a recursive estimator. Martin and Masreliez (1975) pointed out the possibility of constructing recursive M-estimators using a stochastic approximation approach. The classical results for stochastic approximation algorithms can be applied rather straightforwardly to investigate the asymptotic properties of recursive M-estimators when the observations are independent.

The behaviour of recursive M-estimators in dependent situations are less known. The pure location parameter case with m-dependent and strongly regular observations is studied in Holst (1980) and Holst (1984) respectively. For practical use some recursive estimator of scale must be constructed and coupled to the estimator of the location parameter. Recursive scale-estimators which are variants of the median absolute deviation are studied in Holst (1985).

A broader approach to the estimating problem is to construct recursive algorithms based on (1.1). In this paper we prove strong convergence of estimators of the form

$$(1.4)\qquad \begin{cases} \eta_{n+1} = \eta_n + (n+1)^{-1}\tilde{H}_n^{(1)}\tilde{\sigma}_n\psi(\tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)), \\ \sigma_{n+1} = \sigma_n + (n+1)^{-1}\tilde{H}_n^{(2)}\tilde{\sigma}_n\chi(\tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)), \\ \eta_0,\ \sigma_0,\ \tilde{H}_0^{(1)},\ \tilde{H}_0^{(2)} \quad \text{arbitrary and finite,} \end{cases}$$

and mainly we discuss the following choice of $H_n^{(1)}$ and $H_n^{(2)}$:

$$(1.5)\qquad \begin{cases} H_n^{(1)} = \left(n^{-1}\sum_{i=1}^{n}\psi'(\tilde{\sigma}_{i-1}^{-1}(y_i-\eta_{i-1}))\right)^{-1} \\ H_n^{(2)} = \left(n^{-1}\sum_{i=1}^{n}\tilde{\sigma}_{i-1}^{-1}(y_i-\eta_{i-1})\chi'(\tilde{\sigma}_{i-1}^{-1}(y_i-\eta_{i-1}))\right)^{-1} \end{cases}$$

With the notation $\tilde{v}$ we mean v truncated above and below.

We consider the case when the observations $\{y_i\}_1^\infty$ can be described by a strictly stationary process satisfying certain strong mixing conditions. For the analysis we assume that ψ and χ satisfy some regularity conditions. These are introduced in Section 2.

Strong convergence of η_n and σ_n and also of the adaptive sequences $H_n^{(1)}$ and $H_n^{(2)}$ is proved in Section 3.

In Englund, Holst and Ruppert (1987) we prove a strong representation theorem for the estimators. It is possible to derive asymptotic distributions using this theorem together with suitable forms of the Central Limit Theorem. When the observations are a sequence of i.i.d. variables it follows that (η_n, σ_n) has the same asymptotic distribution as the nonrecursive estimator (T_n, S_n). Comments on the asymptotic distribution is given in Section 4. Further we discuss whether our choice of $H_n^{(1)}$ and $H_n^{(2)}$ is optimal or if it is possible to find a better one. We consider a gain matrix which might be preferred, but this matrix contains unknown parameters which like a and b must be estimated, and this leads to an expansion of the dimension of the parameter.

In Section 5 we illustrate the behaviour of the estimates for Huber's Proposal 2 when the observations are i.i.d. with a contaminated normal distribution.

2. NOTATIONS AND ASSUMPTIONS.

To incorporate the adaptive sequences $H_n^{(1)}$ and $H_n^{(2)}$ we rewrite the algorithm in the following way

$$(2.1) \qquad \begin{cases} \theta_{n+1} = \theta_n + (n+1)^{-1} H_n h(\theta_n, y_{n+1}), \\ \theta_0,\ H_0 \ \text{arbitrary and finite,} \end{cases}$$

where

$$\theta_n = (\eta_n, \sigma_n, a_n, b_n)^T.$$

Further

$$h(\theta_n, y_{n+1}) = \begin{pmatrix} \tilde{\sigma}_n \psi(u_{n+1}) \\ \tilde{\sigma}_n \chi(u_{n+1}) \\ \psi'(u_{n+1}) - a_n \\ u_{n+1} \chi'(u_{n+1}) - b_n \end{pmatrix}$$

with

$$u_n = \tilde{\sigma}_{n-1}^{-1}(y_n - \eta_{n-1})$$

and

(2.2) $$H_n = \mathrm{diag}(\tilde{a}_n^{-1}, \tilde{b}_n^{-1}, 1, 1)$$

With the notation $\tilde{\sigma}_n$ we mean σ_n truncated above by a large positive number ν_2 and below by a small positive number ν_1 so that

(2.3) $$\tilde{\sigma}_n = \begin{cases} \nu_1 & \text{if} \quad \tilde{\sigma}_n < \nu_1, \\ \sigma_n & \text{if} \quad \nu_1 \le \tilde{\sigma}_n \le \nu_2, \\ \nu_2 & \text{if} \quad \tilde{\sigma}_n > \nu_2. \end{cases}$$

Throughout the paper it is understood that $\nu_1 \le \sigma \le \nu_2$. The above notation will also be used for $\tilde{a}_n$ and $\tilde{b}_n$. Note that

$$a_n = n^{-1} \sum_{j=1}^{n} \psi'(u_j)$$

and

$$b_n = n^{-1} \sum_{j=1}^{n} u_j \chi'(u_j)$$

so that with H_n defined as in (2.2) we get algorithm (1.4) with $H_n^{(1)}$ and $H_n^{(2)}$ given by (1.5).

Define

$$\bar{h}(x) = E(h(x, y_1))$$

and let θ be the solution of $\bar{h}(\theta) = 0$, where $\theta = (\eta, \sigma, a, b)^T$, that is with $z_1 = \sigma^{-1}(y_1 - \eta)$

(2.4) $$\begin{cases} E(\psi(z_1)) = 0, \\ E(\chi(z_1)) = 0, \\ E(\psi'(z_1)) = a, \\ E(z_1 \chi'(z_1)) = b. \end{cases}$$

Let $F_1^m = F(y_1,\dots,y_m)$ be the σ-algebra generated by the random variables $y_1,\dots,y_m$. The sequence of strong mixing coefficients α_i is defined

$$\alpha_i = \sup_m \alpha(F_1^{m-i}, F_m^\infty) = \sup_m \sup_{F \in F_1^{m-i}, G \in F_m^\infty} |P(FG)-P(F)P(G)| .$$

Further, we need the following notations $n(k) = [k^\delta]$ for some $\delta>2$ and

$$\rho_k = \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} = O(k^{-1}).$$

The constant C is positive and may change from line to line. For shortness we usually write z instead of z_1 below.

Finally we list the following assumptions for later use.

<u>A1.</u> The sequence of observations $\{y_i\}_1^\infty$ is strictly stationary and strong mixing with $\Sigma_{i=1}^\infty \alpha_i^{1-\varepsilon} < \infty$ for some $0 < \varepsilon < 1$. The marginal distribution is symmetric, continuous and positive in a neighbourhood of η.

<u>A2.</u> The function $h(x,y)$ is bounded and Lipschitz-continuous both as a function of x and y i.e.

$$\|h(x_1,y)-h(x_2,y)\| \le K_1 \|x_1-x_2\|$$
$$\|h(x,y_1)-h(x,y_2)\| \le K_2 \|y_1-y_2\|$$

for some positive constants K_1 and K_2.

<u>A3.</u> The function $\psi(\cdot)$ is bounded, increasing (strictly increasing in a neighbourhood of zero) and odd. The function $\chi(\cdot)$ is bounded, increasing on $(0,\infty)$ (strictly increasing in a neighbourhood of zero) and even.

<u>A4.</u> The function $\bar{h}(\cdot)$ satisfies $\bar{h}(\theta) = 0$.

<u>A5.</u> The following functions exist and are bounded:
$\psi^{(k)}(x)$ for $1 \le k \le 3$, $x^k\psi^{(k)}(x)$ for $1 \le k \le 2$, $x^2\psi^{(3)}(x)$,
$\chi^{(k)}(x)$ for $1 \le k \le 2$ and $x^k\chi^{(k)}(x)$ for $1 \le k \le 3$.

Note that A2 holds if the functions $x\psi'(x)$, $x\psi''(x)$, $x\chi'(x)$ and $x^2\chi''(x)$ exist and are bounded and that A5 is a strong assumption which is used in Section 4 only.

3. ALMOST SURE CONVERGENCE.

In this section we study almost sure convergence of the algorithm (2.1). It is proved in Theorem 3.1 that $\theta_n \to \theta$ a.s., where θ solves $\bar{h}(\theta) = 0$. The proof consists of two parts. Following Ruppert (1983) we show that

$$\theta_{n(k+1)} = \theta_{n(k)} + \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} H_i \bar{h}(\theta_{n(k)}) + o(k^{-1}) \,. \tag{3.1}$$

This is accomplished by writing

$$\begin{aligned}
\theta_{n(k+1)} = \theta_{n(k)} &+ \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} H_i \bar{h}(\theta_{n(k)}) + \\
&+ \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} H_{n(k)} (h(\theta_{n(k)}, y_{i+1}) - \bar{h}(\theta_{n(k)})) + \\
&+ \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} (H_i - H_{n(k)}) (h(\theta_{n(k)}, y_{i+1}) - \bar{h}(\theta_{n(k)})) + \\
&+ \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} H_i (h(\theta_i, y_{i+1}) - h(\theta_{n(k)}, y_{i+1})) \\
= \theta_{n(k)} &+ \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} H_i \bar{h}(\theta_{n(k)}) + \\
&+ R_{k,n(k+1)-1} + S_{k,n(k+1)-1} + T_{k,n(k+1)-1},
\end{aligned} \tag{3.2}$$

say, and then we prove that $R_{k,n(k+1)-1}$, $S_{k,n(k+1)-1}$ and $T_{k,n(k+1)-1}$ all are $o(k^{-1})$. The most involved expression, $R_{k,n(k+1)-1}$, is handled in Lemma 3.3, which is a lemma by Ruppert (1983, Lemma 3.2). The second part of the proof is to show that (3.1) is sufficient to establish convergence. This is verified using Lemma 3.4, which is proved by a technique similar to the one used by Blum (1954).

In Lemma 3.1 we prove that $\{g(y_i)\}_1^\infty$ is a mixingale with parameters ψ_m of size $-\frac{1}{2}$ and c_n a constant if $g(\cdot)$ is a bounded function with $E(g(y_1)) = 0$. For a definition of mixingales and notations, see McLeish (1975). Also the result in Lemma 3.2 is a mixingale inequality by McLeish (1975, Theorem 1.6).

<u>LEMMA 3.1</u> Let $g(\cdot)$ be a bounded Borel-measurable function with $E(g(y_i)) = 0$. If A1 holds then $\{g(y_i)\}_1^\infty$ is a mixingale with parameters ψ_m of size $-\frac{1}{2}$ and c_n a constant.

<u>Proof</u> Let $F_m^n = F\{y_m,\ldots,y_n\}$. Lemma 2.1 by McLeish (1975) with $p = 2$ and $r = \infty$ gives

$$\|E(g(Y_n)|F_{-\infty}^{n-m})-E(g(Y_n))\|_2 = \|E(g(Y_n)|F_{-\infty}^{n-m})\|_2$$

$$\le 2(2^{\frac{1}{2}}+1)\sqrt{\alpha(F_{-\infty}^{n-m},F_n^\infty)}\|g(Y_n)\|_\infty$$

$$\le C\alpha_m^{\frac{1}{2}} ,$$

that is $\psi_m = \alpha_m^{\frac{1}{2}}$ and $c_n = C$. The fact that $\Sigma_{i=1}^\infty \alpha_i^{1-\epsilon} < \infty$ for some $0 < \epsilon < 1$ implies that ψ_m is of size $-\frac{1}{2}$ according to McLeish (1975, p. 831). This proves the lemma. □

<u>LEMMA 3.2</u> Let $\{g(y_i)\}_1^\infty$ be defined as in Lemma 3.1. Then there exists a constant c such that

$$E(\max_{n\le m}|\sum_{i=1}^n d_i g(y_i)|^2) \le c\sum_{i=1}^m d_i^2$$

for all m and constants $d_1,\ldots,d_m$.

<u>Proof</u> It is obvious from Lemma 3.1 that $\{d_i g(y_i)\}_1^\infty$ is a mixingale with parameters ψ_m of size $-\frac{1}{2}$ and $c_n = d_n C$. Theorem 1.6 by McLeish (1975) proves the lemma. □

LEMMA 3.3 If A1-A2 hold then

$$\sup_{x\in R^4} \max_{n(k)\le \ell < n(k+1)} \rho_k^{-1} \left\| \sum_{i=n(k)}^{\ell} i^{-1}(h(x,y_{i+1})-\bar{h}(x)) \right\| \to 0$$

when $k \to \infty$.

Proof According to Ruppert (1983, Lemma 3.2), we have to verify his assumptions A3-A6. A3 is obvious, A4 is exactly Lemma 3.2 above if $r = \infty$ and A5 is satisfied since $h(x,y)$ is bounded. Finally A6 follows from the Lipschitz continuity of $h(x,y)$ as a function of y. This proves Lemma 3.3. □

LEMMA 3.4 Let $t(\cdot)$ be a bounded function from R^n to R, $x^{(j)}$ an element of R^1 and $\{x_k\}$ a sequence of r.v. satisfying the following assumptions:

B1. $x_{k+1}^{(j)} - x_k^{(j)} = D_k t(x_k) + o(k^{-1})$ a.s. for some positive sequence $\{D_k\}_1^\infty$ satisfying $K_1 k^{-1} \le D_k \le K_2 k^{-1}$ where K_1 and K_2 are positive constants.

B2. For all $\gamma > 0$ there are $\delta_1, \delta_2 > 0$ and N_γ satisfying $\sup t(x_k) = -\delta_1$, where the supremum is for $\{x_k : x_k^{(j)} > x^{(j)}+\gamma\}$, and

$\inf t(x_k) = \delta_2$, where the infimum is for $\{x_k : x_k^{(j)} < x^{(j)}-\gamma\}$, for all $k \ge N_\gamma$.

Then $x_k^{(j)} \to x^{(j)}$ a.s.

Proof Assume that $x_k^{(j)} \to \infty$. The assumptions make it possible to find a constant N_1 such that $D_k t(x_k) + o(k^{-1}) < 0$ for $k > N_1$, and thus $x_{k+1}^{(j)} < x_k^{(j)}$ and hence we get a contradiction. (The case $x_k^{(j)} \to -\infty$ is treated in the same way.) Now assume that $x_k^{(j)}$ doesn't converge, that is $\liminf x_k^{(j)} < \limsup x_k^{(j)}$, and also assume that $\limsup x_k^{(j)} > x^{(j)}$. (The case $\limsup x_k^{(j)} \le x^{(j)}$ is handled by a similar argument.) Define γ from the relation $\limsup x_k^{(j)} = x^{(j)} + 3\gamma$. Take N_2 so large that $-D_k\delta_1 + o(k^{-1}) < 0$ for $k > N_2$. Then we can find $N_2 \le n$, $m > n+1$

such that $x^{(j)} < x_n^{(j)} < x^{(j)} + \gamma$, $x^{(j)} + \gamma \le x_k^{(j)} \le x^{(j)} + 2\gamma$ for $k = n+1,\dots,m-1$ and $x_m^{(j)} > x^{(j)} + 2\gamma$. This is possible since $x_{k+1}^{(j)} - x_k^{(j)} \to 0$. Now

$$x_m^{(j)} - x_n^{(j)} = \sum_{k=n}^{m-1} (D_k t(x_k) + o(k^{-1})) < D_n t(x_n) + o(n^{-1})$$

and this quantity can be made arbitrarily small, which is a contradiction. □

Theorem 3.1 will now be stated.

<u>THEOREM 3.1</u> Let θ_n be generated by algorithm (2.1). If A1-A4 hold, then $\theta_n \to \theta$ a.s. as $n \to \infty$.

<u>Proof</u> The first part is to prove that

$$\theta_{n(k+1)} = \theta_{n(k)} + \sum_{i=n(k)}^{n(k+1)-1} (i+1)^{-1} H_i \bar{h}(\theta_{n(k)}) + o(k^{-1}) . \tag{3.3}$$

For $n(k) \le \ell < n(k+1)$ we have

$$\begin{aligned}
\theta_{\ell+1} &= \theta_\ell + (\ell+1)^{-1} H_\ell h(\theta_\ell, y_{\ell+1}) \\
&= \theta_{n(k)} + \sum_{i=n(k)}^{\ell} (i+1)^{-1} H_i h(\theta_i, y_{i+1}) \\
&= \theta_{n(k)} + \sum_{i=n(k)}^{\ell} (i+1)^{-1} H_i \bar{h}(\theta_{n(k)}) + \\
&\quad + \sum_{i=n(k)}^{\ell} (i+1)^{-1} H_{n(k)} (h(\theta_{n(k)}, y_{i+1}) - \bar{h}(\theta_{n(k)})) + \\
&\quad + \sum_{i=n(k)}^{\ell} (i+1)^{-1} (H_i - H_{n(k)}) (h(\theta_{n(k)}, y_{i+1}) - \bar{h}(\theta_{n(k)})) + \\
&\quad + \sum_{i=n(k)}^{\ell} (i+1)^{-1} H_i (h(\theta_i, y_{i+1}) - h(\theta_{n(k)}, y_{i+1})) \\
&= \theta_{n(k)} + \sum_{i=n(k)}^{\ell} (i+1)^{-1} H_i \bar{h}(\theta_{n(k)}) + R_{k,\ell} + S_{k,\ell} + T_{k,\ell} .
\end{aligned}$$

The fact that $\|R_{k,\ell}\| = o(k^{-1})$ follows from Lemma 3.3 and due to the boundedness of $h(x,y)$ we also have $\|S_{k,\ell}\| = o(k^{-1})$ if we can prove that

$$\sup_{n(k)<i<n(k+1)} \|H_i - H_{n(k)}\| = o(1) .$$

This follows easily for our choice of H_i. The term $T_{k,\ell}$ is treated by writing

$$\|T_{k,\ell}\| = \|\sum_{i=n(k)}^{\ell} (i+1)^{-1} H_i (h(\theta_i, y_{i+1}) - h(\theta_{n(k)}, y_{i+1}))\|$$

$$\le C \sum_{i=n(k)}^{\ell} (i+1)^{-1} \|h(\theta_i, y_{i+1}) - h(\theta_{n(k)}, y_{i+1})\|$$

$$\le C \sum_{i=n(k)}^{\ell} (i+1)^{-1} \|\theta_i - \theta_{n(k)}\|$$

$$\le C\rho_k \max_{n(k)\le i<n(k+1)} \|\theta_i - \theta_{n(k)}\|$$

since $h(x,y)$ is Lipschitz continous in x. Now we have proved that

$$\|\theta_{\ell+1} - \theta_{n(k)}\| \le C_1\rho_k + o(\rho_k) + C_2\rho_k \max_{n(k)\le i<n(k+1)} \|\theta_i - \theta_{n(k)}\| .$$

The inequality

$$\max_{n(k)\le i<n(k+1)} \|\theta_i - \theta_{n(k)}\| \le (C_1\rho_k + o(\rho_k))/(1 - C_2\rho_k) \tag{3.4}$$

for large k gives $\|T_{k,\ell}\| = o(k^{-1})$. Summarizing we have verified (3.3).

The second part of the proof is to show that this gives the result stated in the theorem. We apply Lemma 3.4 to the components of the vector $\bar{h}(\theta_{n(k)})$. It is obvious that B1 holds and it remains to verify B2 for all components. We start at $\bar{h}^{(1)}(\theta_{n(k)})$ and take the components in order. The convergence of $\eta_{n(k)}$ follows because $\bar{h}^{(1)}(\theta_{n(k)})$ satisfies B2 since $\psi(\cdot)$ is increasing and odd. For $\sigma_{n(k)}$ we have

$$\bar{h}^{(2)}(\theta_{n(k)}) = \bar{h}^{(2)}(\theta_{n(k)}) - \bar{h}^{(2)}(\kappa_{n(k)}) + \bar{h}^{(2)}(\kappa_{n(k)}) ,$$

where

$$\kappa_{n(k)} = (\eta, \sigma_{n(k)}, a_{n(k)}, b_{n(k)})^T .$$

Assumptions A1 and A4 implies that $\bar{h}^{(2)}(\kappa_{n(k)}) = -\delta$ if $\sigma_{n(k)} > \sigma+\gamma$ because $\chi(\cdot)$ is increasing on $(0,\infty)$ and even and the first part of the assumption is satisfied from the fact that

$$\|\bar{h}^{(2)}(\theta_{n(k)}) - \bar{h}^{(2)}(\kappa_{n(k)})\| \le C\|\eta_{n(k)}-\eta\| \le \delta/2$$

if k is large enough. This is due to the proved part above and the Lipschitz-continuity of $\bar{h}(x)$ as a function of x. The second part of the assumption follows in the same way. The convergence of $a_{n(k)}$ follows because

$$\begin{aligned}\bar{h}^{(3)}(\theta_{n(k)}) &= \bar{h}^{(3)}(\theta_{n(k)}) - \bar{h}^{(3)}(\lambda_{n(k)}) + \bar{h}^{(3)}(\lambda_{n(k)}) \\ &= \bar{h}^{(3)}(\theta_{n(k)}) - \bar{h}^{(3)}(\lambda_{n(k)}) + a - a_{n(k)}\end{aligned}$$

where

$$\lambda_{n(k)} = (\eta,\sigma,a_{n(k)},b_{n(k)})^T .$$

The Lipschitz-continuity makes it possible to choose N such that

$$\|\bar{h}^{(3)}(\theta_{n(k)}) - \bar{h}^{(3)}(\lambda_{n(k)})\| \le C(\|\eta-\eta_{n(k)}\|+\|\sigma-\sigma_{n(k)}\|) < \delta/2$$

for all $k > N$, and this proves that $a_{n(k)} \to a$. A similar argument shows that $b_{n(k)} \to b$.

The relation $\|\theta_{n(k+1)}-\theta_{n(k)}\| \to 0$ and the previous result (3.4) proves the remaining part of the theorem. □

4. COMMENTS ON THE ASYMPTOTIC DISTRIBUTION AND ON THE CHOICE OF THE ADAPTIVE MATRIX.

In Section 3 we proved strong consistency of the algorithm (2.1). In order to discuss our choice of H_n we also need results for the asymptotic distribution of the algorithm.

The asymptotic distribution can be derived from a strong representation theorem which is proved in Englund, Holst and Ruppert (1987). The same theorem is stated here without proof to facilitate the discussion below. (By the notation $\tilde{x}$ in this section we mean a continuous and differentiable version of (2.3).)

<u>THEOREM 4.1</u> If A1, A3-A5 hold and θ_n is given by the algorithm (2.1), then there exists $\varepsilon > 0$ such that

$$n^{\frac{1}{2}}(\theta_n-\theta) = n^{-\frac{1}{2}} \begin{pmatrix} \sum_{k=1}^{n} a^{-1}\sigma\psi(z_k) \\ \sum_{k=1}^{n} b^{-1}\sigma\chi(z_k) \\ d_1\sum_{k=1}^{n} \log(\frac{k}{n})\chi(z_k) + \sum_{k=1}^{n} (\psi'(z_k)-a) \\ d_2\sum_{k=1}^{n} \log(\frac{k}{n})\chi(z_k) + \sum_{k=1}^{n} (z_k\chi'(z_k)-b) \end{pmatrix} + O(n^{-\varepsilon}),$$

where $z_k = \sigma^{-1}(y_k-\eta)$, $d_1 = b^{-1}E(z_1\psi''(z_1))$ and $d_2 = 1+b^{-1}E(z_1^2\chi''(z_1))$.

For a sequence of *independent* observations the theorem gives

$$n^{\frac{1}{2}}(\theta_n-\theta) \in \text{As } N(0,V),$$

where

$$V = \begin{pmatrix} V_{11} & 0 & 0 & 0 \\ & V_{22} & V_{23} & V_{24} \\ & & V_{33} & V_{34} \\ & & & V_{44} \end{pmatrix}$$

with variance elements

$$V_{11} = a^{-2}\sigma^2E\psi^2(z),$$
$$V_{22} = b^{-2}\sigma^2E\chi^2(z),$$
$$V_{33} = V(\psi'(z)) + 2d_1^2V(\chi(z)) - 2d_1C(\chi(z),\psi'(z)),$$

$$V_{44} = V(z\chi'(z)) + 2d_2^2V(\chi(z)) - 2d_2C(\chi(z), z\chi'(z)),$$

and covariance elements

$$V_{23} = -d_1b^{-1}\sigma V(\chi(z)) + b^{-1}\sigma C(\chi(z), \psi'(z)),$$
$$V_{24} = -d_2b^{-1}\sigma V(\chi(z)) + b^{-1}\sigma C(\chi(z), z\chi'(z)),$$
$$V_{34} = 2d_1d_2V(\chi(z)) - d_1C(\chi(z), z\chi'(z)) - d_2C(\chi(z), \psi'(z)) + \\ + C(z\chi'(z), \psi'(z)).$$

In the remaining part of this section we discuss whether H_n in Section 3 is optimal or if we can find a better one.

Given a recursive algorithm it is well known that the "optimal" adaptive matrix H^{opt} is the negative inverse of the derivates of $\bar{h}(\theta)$. Here we get

$$H^{opt} = \begin{pmatrix} E(\psi'(z)) & E(z\psi'(z)-\psi(z)) & 0 & 0 \\ E(\chi'(z)) & E(z\chi'(z)-\chi(z)) & 0 & 0 \\ E(\sigma^{-1}\psi''(z)) & E(\sigma^{-1}z\psi''(z)) & 1 & 0 \\ E(\sigma^{-1}(\chi'(z)+z\chi''(z))) & E(\sigma^{-1}(z\chi'(z)+z^2\chi''(z))) & 0 & 1 \end{pmatrix}^{-1}$$

If ψ is odd and χ even this reduces to

$$H^{opt} = \begin{pmatrix} a^{-1} & 0 & 0 & 0 \\ 0 & b^{-1} & 0 & 0 \\ 0 & -(b\sigma)^{-1}E(z\psi''(z)) & 1 & 0 \\ 0 & -\sigma^{-1}(1+b^{-1}E(z^2\chi''(z))) & 0 & 1 \end{pmatrix}, \tag{4.1}$$

where as above $a = E(\psi'(z))$ and $b = E(z\chi'(z))$. The values of a, b, $E(z\psi''(z))$ and $E(z^2\chi''(z))$ are in general unknown and if we try to estimate $E(z\psi''(z))$ and $E(z^2\chi''(z))$ we get more elements in the parameter vector.

It is however worth noting that this more complicated algorithm may reduce the asymptotic variances of a_n and b_n. If we assume that we know the values of $d_1 = b^{-1}E(z\psi''(z))$ and $d_2 = 1+b^{-1}E(z^2\chi''(z))$ and insert them and the truncated estimates of a and b in the matrix H^{opt} we get the algorithm

$$(4.2)\qquad \begin{cases} \eta'_{n+1} = \eta'_n + (n+1)^{-1}\tilde{\sigma}'_n\psi(u'_{n+1})/\tilde{a}'_n, \\ \sigma'_{n+1} = \sigma'_n + (n+1)^{-1}\tilde{\sigma}'_n\chi(u'_{n+1})/\tilde{b}'_n, \\ a'_{n+1} = a'_n + (n+1)^{-1}(\psi'(u'_{n+1}) - d_1\chi(u'_{n+1}) - a'_n), \\ b'_{n+1} = b'_n + (n+1)^{-1}(u'_{n+1}\chi'(u'_{n+1}) - d_2\chi(u'_{n+1}) - b'_n), \end{cases}$$

where $u'_{n+1} = (y_{n+1}-\eta'_n)/\tilde{\sigma}'_n$. It is easy to prove that this algorithm satisfies $\theta'_n \to \theta$ a.s. and from the technique used in Englund, Holst and Ruppert (1987) it also follows for *independent* observations that

$$n^{1/2}(\theta'_n-\theta) \in \mathrm{As}\, N(0,V'),$$

where

$$V' = \begin{pmatrix} V_{11} & 0 & 0 & 0 \\ & V_{22} & V_{23} & V_{24} \\ & & V'_{33} & V'_{34} \\ & & & V'_{44} \end{pmatrix}$$

The only difference between V and V' is the elements

$$V'_{33} = V(\psi'(z)) + d_1^2V(\chi(z)) - 2d_1C(\chi(z),\psi'(z))$$

$$V'_{44} = V(z\chi'(z)) + d_2^2V(\chi(z)) - 2d_2C(\chi(z),z\chi'(z)).$$

$$V'_{34} = d_1d_2V(\chi(z)) - d_1C(\chi(z),z\chi'(z)) - d_2C(\chi(z),\psi'(z)) + \\ + C(z\chi'(z),\psi'(z)).$$

Note that the variances V_{33} and V_{44} for the algorithm in Section 2 both are larger than V'_{33} and V'_{44}.

As an example we take Huber's Proposal 2 with $k = 1.5$. Although the functions defined in (1.2) do not satisfy A2 and A5 it is conjectured in Englund, Holst and Ruppert (1987) that the theorem is valid if d_1 and d_2 are interpreted as $d_1 = -2b^{-1}kf(k)$ and $d_2 = 2 - 4b^{-1}k^3f(k)$, where f is the density of z. For this choice and independent $N(0,1)$ distributed random variables we get

$$V = \begin{pmatrix} 1.0371 & 0 & 0 & 0 \\ & 0.6894 & 0.0621 & 0.2797 \\ & & 0.1641 & 0.2262 \\ & & & 1.2326 \end{pmatrix}$$

and

$$V' = \begin{pmatrix} 1.0371 & 0 & 0 & 0 \\ & 0.6894 & 0.0621 & 0.2797 \\ & & 0.0600 & 0.2699 \\ & & & 1.2143 \end{pmatrix}.$$

Observe that $b_n' = 2k^2a_n' - 2(k^2 - \beta_k)$ for Huber's Proposal 2, which implies that $\rho(a_n',b_n') = 1$ and hence the number of components of θ_n' reduces to three.

It is our intention to study algorithm (4.2) with estimates of d_1 and d_2 in the near future. For Huber's Proposal 2 we only have to estimate d_1 and this makes use of H^{opt} more feasible.

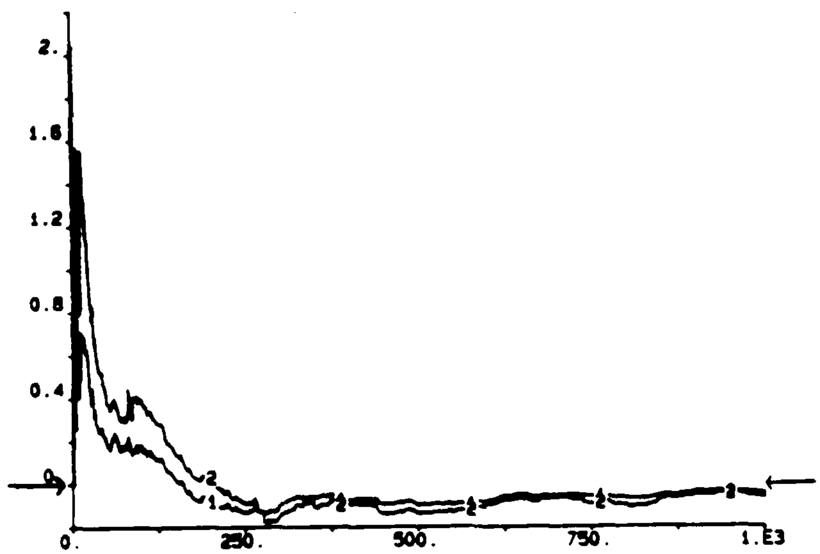
5. A NUMERICAL EXAMPLE.

In this section we give a numerical example of the adaptive estimator defined in Section 2 when $\{y_t\}_1^{1000}$ is a sequence of independent r.v. with a contaminated normal distribution $0.9N(0,1)+0.1N(0,25)$. We will use Huber's Proposal 2, defined in (1.2). The constant k is chosen to 1.5 which makes $\beta_{1.5} = 0.7784$. The variables σ_n, a_n and b_n are all truncated below by 0.1 and above by 10. To avoid that bad early estimates of $\tilde{\sigma}_n$, $\tilde{a}_n$ and $\tilde{b}_n$ influence the results too much we take $H_n = I$ and

$$h(\theta_n, y_{n+1}) = \begin{pmatrix} \psi(\tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)) \\ \chi(\tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)) \\ \psi'(\tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)) - a_n \\ \tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)\chi'(\tilde{\sigma}_n^{-1}(y_{n+1}-\eta_n)) - b_n \end{pmatrix}$$

if $n \leq 50$. The initial value is $\theta_0 = (0,1,0,0)^T$ and the solution of (2.4) is $(\eta,\sigma,a,b)^T = (0,\ 1.1346,\ 0.8468,\ 0.8024)^T$.

The figures below are produced to give an impression of the behaviour of the recursive estimates. The performance of η_n, σ_n, a_n and b_n for $n = 1,\ldots,1000$ is shown in Figures 5.1 - 5.4 respectively. Also the recursive least squares estimator of η, the sample mean, is given in Figure 5.1 for comparison. The arrows in the figures indicate the convergence points.



<u>Fig. 5.1.</u> 1: η_n

2: sample mean

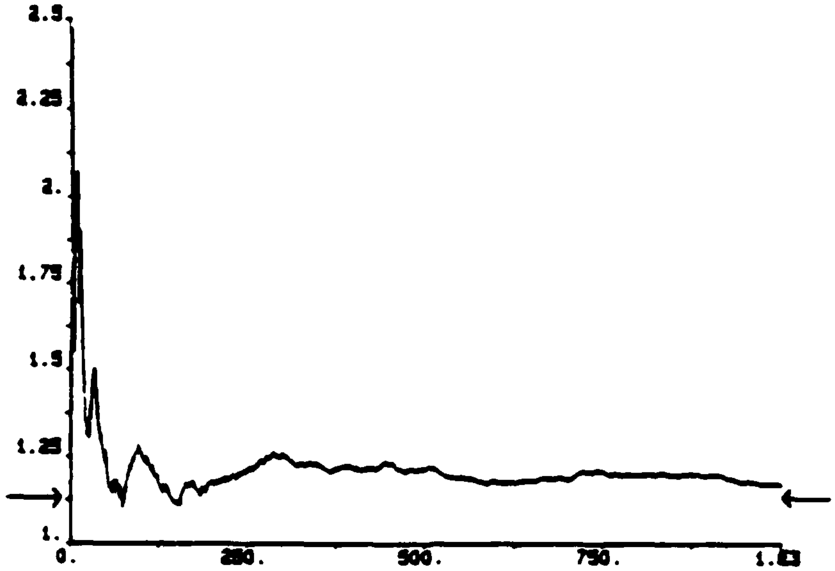


Fig. 5.2. σ_n

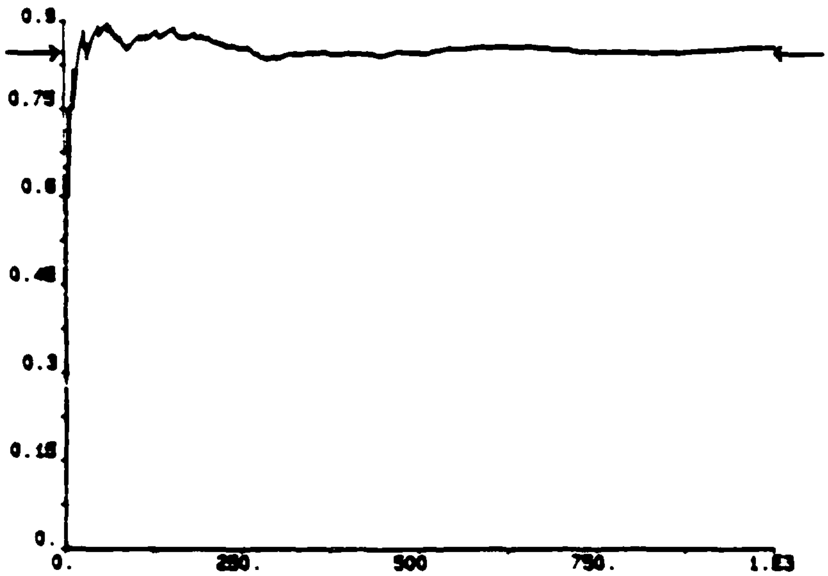


Fig. 5.3. a_n

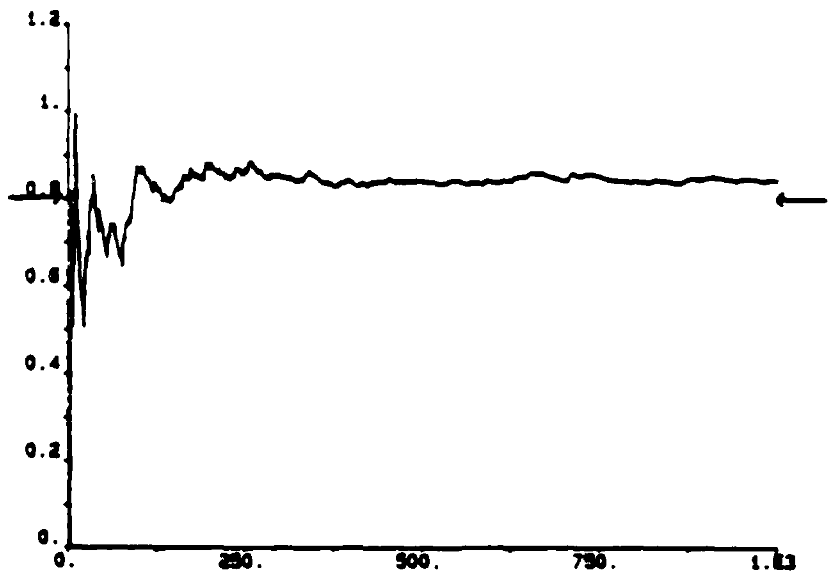


Fig. 5.4. b_n

Finally we mention that the asymptotic variance is 1.3977 for the recursive estimator, while the least squares estimator has the asymptotic variance 3.4000.

6. REFERENCES.

Blum, J. (1954). Approximation methods which converge with probability one. Ann. Math. Statist. 25, 382-386.

Englund, J-E., Holst, U., and Ruppert, D. (1987). A representation theorem for generalized Robbins-Monro processes and applications. Univ. of Lund, Stat. Research Report 1987:1.

Hampel et al. (1986). Robust statistics, the approach based on influence functions. Wiley-Interscience.

Holst, U. (1980). Convergence of a recursive stochastic algorithm with m-dependent observations. Scand. J. Statist. 7, 207-215.

Holst, U. (1984). Convergence of a recursive robust algorithm with strongly regular observations. Stochastic Process. Appl. 16, 305-320.

Holst, U. (1985). Recursive M-estimators of location. Manuscript submitted to Comm. in Statistics.

Huber, P (1981). Robust statistics. Wiley-Interscience.

Martin, R.D. and Masreliez, C.J. (1975). Robust estimation via stochastic approximation. IEEE Trans. Inform. Theory 21, 263-271.

McLeish, D.L. (1975). A maximal inequality and dependent strong laws. Ann. of Prob. 3, 829-839.

Ruppert, D (1983). Convergence of stochastic approximation algorithms with non-additive dependent disturbances and applications. Lecture Notes in Statistics 20. Springer-Verlag.

END

DATE

FILMED

DEC.

1987